

Linguistic metrics implemented in VizLing a visual linguistic analytics tool

Thomas Gaillat

Anas Knefati

Antoine Lafontaine

Abstract

This document lists the metrics implemented in the VizLing application. ViZling is a tool for the visualisation of linguistic complexity in second language (L2) learner writings. We show how learner writings in English can be exploited with Natural Language Processing (NLP) tools to compute and visualise complexity metrics. An annotated corpus in terms of proficiency is used as a benchmark for the comparison with new learner metrics. These visualisations are used as feedback for learners whose metrics are superimposed with proficiency levels.

Metrics, dimension and tools employed in the VizLing pipeline

1. Syntactic complexity implemented in L2SCA (Table remodeled from Lu 2010)

Metrics	Dimension
Mean Length of Clause (MLC) Mean Length of Sentence (MLS) Mean Length of T-Unit (MLT)	length of production unit
Sentence Complexity ratio (C.S)	sentence complexity
T-Unit Complexity ratio Complex T-Unit ratio (CT/T) Dependent Clause ratio (DC/C) Dependent Clauses per T-unit (DC/T)	subordination
Coordinate Phrases per Clause (CP/C) Coordinate Phrases per T-unit (CP/T) Sentence coordination ratio (T.S)	coordination
Complex Nominals per Clause (CN/C) Complex Nominals per T-unit (CN/T)	nominal structures
Verb Phrases per T-unit (VP/T)	verb structures

2. Lexical richness using the Quanteda R-package with the function `textstat_lexdiv()`

Metrics	Dimension
Type Token Ratio (TTR 5 versions) Uber Index	density
Yule's K Simpson's D Herdan's Vm (type repetitions)	diversity (type repetitions)
Maas indices (3 versions)	diversity (type variation)

3. Readability using the Quanteda R-package with the function `textstat_readability()`, the metrics used are:

- Automated Readability Index(ARI 3 versions)
- Bormuth's Mean Cloze Formula (MC)
- Bormuth's Grade Placement score (PC)
- Coleman (2 versions)
- Coleman-Liau (3 versions)
- Dale-Chall (3 versions)
- Dale-Chall (3 versions)
- Danielson-Bryan (2 versions)
- Degrees of Reading Power (DRP)
- Easy Listening Formula (ELF)
- Farr.Jenkins.Paterson
- Flesch (3 versions)
- Gunning's Fog Index (3 versions)
- Forecast (2 versions)
- Fucks
- Linsear.Write
- LIW
- Neue Wiener Sachtextformeln(nWS 4 versions)
- Anderson's Readability Index (RIX)
- Scrabble
- Simple Measure of Gobbledygook(SMOG 4 versions)
- Spache (2 versions)
- Traenkle.Bailer (2 versions)
- Strain
- Wheeler.Smith
- meanSentenceLength
- meanWordSyllables